American Psychological Association

Emotion

© 2020 American Psychological Association
ISSN: 1528-3542

2022, Vol. 22, No. 5, 894–906
http://dx.doi.org/10.1037/emo0000770

Enhanced Recognition of Vocal Emotions in Individuals With Naturally Good Musical Abilities

Ana Isabel Correia
Instituto Universitário de Lisboa (ISCTE-IUL)

São Luís Castro
University of Porto

Chloe MacGregor and Daniel Müllensiefen
Goldsmiths, University of London

E. Glenn Schellenberg
University of Toronto Mississauga

César F. Lima
Instituto Universitário de Lisboa (ISCTE-IUL) and University College London

Music training is widely assumed to enhance several nonmusical abilities, including speech perception, executive functions, reading, and emotion recognition. This assumption is based primarily on cross-sectional comparisons between musicians and nonmusicians. It remains unclear, however, whether training itself is necessary to explain the musician advantages, or whether factors such as innate predispositions and informal musical experience could produce similar effects. Here, we sought to clarify this issue by examining the association between music training, music perception abilities and vocal emotion recognition. The sample (N = 169) comprised musically trained and untrained listeners who varied widely in their musical skills, as assessed through self-report and performance-based measures. The emotion recognition tasks required listeners to categorize emotions in nonverbal vocalizations (e.g., laughter, crying) and in speech prosody. Music training was associated positively with emotion recognition across tasks, but the effect was small. We also found a positive association between music perception abilities and emotion recognition in the entire sample, even with music training held constant. In fact, untrained participants with good musical abilities were as good as highly trained musicians at recognizing vocal emotions. Moreover, the association between music training and emotion recognition was fully mediated by auditory and music perception skills. Thus, in the absence of formal music training, individuals who were "naturally" musical showed musician-like performance at recognizing vocal emotions. These findings highlight an important role for factors other than music training (e.g., predispositions and informal musical experience) in associations between musical and nonmusical domains.

Keywords: emotion, music, training, aptitude, voice

Much attention has been devoted to the possibility of associations between musical experience and nonmusical abilities, including speech perception (e.g., Coffey, Chepesiuk, Herholz, Baillet, & Zatorre, 2017), reading ability (e.g., Moreno et al., 2009; Swaminathan, Schellenberg, & Venkatesan, 2018), phonological awareness (e.g., Degé & Schwarzer, 2011; Moritz, Yampolsky, Papadelis, Thomson, & Wolf, 2012), working memory (e.g., Roden, Grube, Bongard, & Kreutz, 2014), executive functions (Moreno et al., 2011; Slevc, Davey, Buschkuehl, & Jaeggi, 2016), IQ (e.g., Schellenberg, 2011), and the perception of speech prosody (e.g., Marques, Moreno, Castro, & Besson, 2007; Schön, Magne, & Besson, 2004). It is often assumed that music training has the power to improve these abilities (e.g., Herholz & Zatorre, 2012; Kraus & Chandrasekaran, 2010; Moreno & Bidelman, 2014; Patel, 2014). In other words, changes in a musician's brain and behavior would have *far transfer* effects, such that they lead to better performance on tasks that are not related to music.

Evidence for far transfer comes primarily from cross-sectional comparisons between classically trained musicians and individuals without training, so-called nonmusicians (e.g., Lima & Castro, 2011; Pinheiro, Vasconcelos, Dias, Arrais, & Gonçalves, 2015; Schön et al., 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007). It also comes from a few longitudinal studies, though evidence from these tends to be weak and more likely to be associated with suboptimal designs (e.g., Sala & Gobet, 2017a, 2017b). Longitu-

This article was published Online First July 27, 2020.

Ana Isabel Correia, Department of Social and Organizational Psychology, Instituto Universitário de Lisboa (ISCTE-IUL); São Luís Castro, Faculty of Psychology and Education Sciences, Universidade do Porto; Chloe MacGregor and Daniel Müllensiefen, Department of Psychology, Goldsmiths, University of London; E. Glenn Schellenberg, Department of Psychology, University of Toronto Mississauga; César F. Lima, Department of Social and Organizational Psychology, Instituto Universitário de Lisboa (ISCTE-IUL) and Institute of Cognitive Neuroscience, University College London.

Funded by grants from Portuguese Foundation for Science and Technology (FCT) awarded to Cesar F. Lima (PTDC/PSI-GER/28274/2017 and IF/00172/2015) and to the Center for Psychology at University of Porto (UID/PSI/00050/2013 and PTDC/PSI-GER/28274/2017).

Correspondence concerning this article should be addressed to César F. Lima, Department of Social and Organizational Psychology, Instituto Universitário de Lisboa (ISCTE-IUL), Avenida das Forças Armadas, 1649-026 Lisboa, Portugal. E-mail: cesar.lima@iscte-iul.pt

dinal studies are typically conducted with children in educational contexts, in which music training programs are compared against no training or training in other domains, such as visual arts (Habibi, Cahn, Damasio, & Damasio, 2016; Moreno et al., 2009, 2011; Schellenberg, Corrigall, Dys, & Malti, 2015; Thompson, Schellenberg, & Husain, 2004).

Studying associations between music training and nonmusical variables helps to inform debates on learning and development, and to improve our understanding of how music relates to other domains at behavioral, cognitive, and brain levels (e.g., Patel, 2014; Peretz & Coltheart, 2003). The primary focus on music *training* reflects a narrow view of musicality, however, because musical skills are diverse and determined by multiple factors other than formal lessons. For example, sophisticated musical abilities can be seen in individuals without any training, and such abilities must be a consequence of informal engagement with music and/or musical predispositions (Bigand & Poulin-Charronnat, 2006; Mankel & Bidelman, 2018; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014; Swaminathan & Schellenberg, 2018). Indeed, recent perspectives on musicality consider a broad range of musical behaviors and skills beyond playing an instrument or taking classes (e.g., informal listening experience, functional uses of music in everyday life, singing along with tunes; Honing, 2019; Krishnan et al., 2018; Müllensiefen, Gingras, Musil, & Stewart, 2014).

Factors other than formal instruction could, therefore, account for the musician advantages reported in the literature. Enhanced capacities of trained individuals might be induced by training, but they could also reflect genetic variables, early informal engagement with music, or facets of musical experience unrelated to formal training per se (as well as more general cognitive, socioeconomic or personality variables; e.g., Swaminathan & Schellenberg, 2018). To distinguish between training itself and these other factors, it is important to study the musical abilities of nonmusicians, and to identify individuals with good abilities despite not being trained. Recent evidence indicates that good music perception skills are associated with good performance in nonmusical domains, regardless of training. For example, such individuals exhibit enhanced phoneme perception in a foreign language (Swaminathan & Schellenberg, 2017) and more efficient neural encoding of speech (Mankel & Bidelman, 2018), mirroring the benefits observed in trained musicians. In short, formal training might not be necessary, or at least not the only factor accounting for the musician advantages in nonmusical domains.

In the present study, we focused on the association between music and one aspect of socioemotional processing, namely the recognition of emotions in vocal expressions. Although music and musical activities are fundamentally linked to socioemotional processes (e.g., Clark, Downey, & Warren, 2015; Koelsch, 2014), this topic remains much less explored than associations with domains such as executive functions (e.g., Moreno et al., 2011; Slevc et al., 2016) or speech processing (e.g., Madsen, Marschall, Dau, & Oxenham, 2019; Mankel & Bidelman, 2018). Some evidence indicates that trained musicians outperform untrained individuals in their ability to recognize emotions in speech prosody, that is, emotional states expressed through a speaker's use of pitch, loudness, timing, and timbre cues in speech (Lima & Castro, 2011; Thompson et al., 2004). Other evidence documents that music training predicts efficient low-level neural encoding (auditory brain stem responses) of purely nonverbal vocalizations such as crying (Strait, Kraus, Skoe, & Ashley, 2009). Neurocognitive pathways for processing music and vocal emotions may overlap, such that formal training in music improves vocal emotional communication in typical and atypical populations (Good et al., 2017).

One possible mechanism is that music training fine-tunes auditory-perceptual abilities that are useful for sensory aspects of voice perception (e.g., pitch and temporal processing). Patel's OPERA hypothesis (Patel, 2011, 2014) uses this view as an explanation for transfer effects from music to speech processing. Another possibility is that because social-emotional interactions are a central component of many musical activities, higher-order aspects of vocal emotional processing are improved by training because the code for music and vocal emotions is at least partly shared (Juslin & Laukka, 2003; see also Clark et al., 2015; Koelsch, 2014; Koelsch, 2015; Pinheiro et al., 2015). Nevertheless, a musician advantage in emotion recognition is not always evident (Park et al., 2015; Trimmer & Cuddy, 2008), and this issue is typically explored in cross-sectional studies that do not take into account individual differences in musical abilities, particularly in nonmusicians. Therefore, it remains unclear whether training itself is necessary to drive the putative advantage, or whether musical predispositions and informal engagement with music could produce similar effects.

Here, our sample of listeners included highly trained musicians and a large number of individuals with minimal or no music training, who were assessed in detail about their music perception abilities, behaviors, and experiences. Our goals were to determine if the advantage for musicians in vocal emotion recognition could be replicated, and to examine the potential role of "natural" individual differences in musical abilities. Specifically, we asked whether having good listening skills, as identified in musical and nonmusical tasks, could also predict the ability to recognize vocal emotions, regardless of music training. In other words, could musically adept individuals with no training approach performance levels similar to those of musicians?

Musical skills, behaviors, and experience were assessed using the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014; Portuguese version, Lima, Correia, Müllensiefen, & Castro, 2018). The Gold-MSI is a self-report tool designed to evaluate music training, music perception abilities, active engagement with music, singing abilities, and emotional responses to music in the general population. We also included performance-based auditory and music perception tasks, which indexed pitch discrimination, duration discrimination, beat perception, and melodic memory. Our outcome measures focused on two sources of nonverbal emotional information in the human voice (e.g., Brück, Kreifelts, & Wildgruber, 2011; Scott, Sauter, & McGettigan, 2010). One was the ability to decode emotions conveyed through prosody in actual speech; the other was the ability to decode emotions conveyed by nonverbal vocalizations (e.g., laughter, crying). Although both prosody and nonverbal vocalizations are vocal signals, their underlying production and perception mechanisms are partly distinct (Pell et al., 2015; Scott et al., 2010). Combining them within the same design allowed us to determine whether associations with music reflect an effect that is specific to prosody, or an effect that extends to the recognition of vocal emotions more generally. Previous research in the area has mostly relied on prosodic

stimuli (Lima & Castro, 2011; Park et al., 2015; Pinheiro et al., 2015; Thompson et al., 2004; Trimmer & Cuddy, 2008).

We predicted that music training would be associated positively with vocal emotion recognition, both for prosody and nonverbal vocalizations, which would represent a replication and extension of previous findings (Lima & Castro, 2011; Thompson et al., 2004). We also expected that auditory and music perception skills would be positively correlated with the ability to recognize vocal emotions, even after accounting for music training. This hypothesis was based on evidence of enhancements in phoneme perception and speech processing in untrained individuals with good music perception skills (Mankel & Bidelman, 2018; Swaminathan & Schellenberg, 2017). Because domain-general cognitive abilities predict both music training *and* music perception skills (Swaminathan & Schellenberg, 2018), a digit span task was included to examine whether observed associations were simply a byproduct of general factors.

More exploratory questions asked whether the link between music and emotion recognition is specific to audition. Lima et al. (2016) identified that individuals with congenital amusia (i.e., a music disorder present throughout development) have deficits in identifying emotions expressed vocally *and* visually through facial expressions. Nevertheless, the role of individual differences in musical abilities among typically developing individuals remains unknown. We also asked whether other aspects of musical expertise and experience (i.e., active engagement with music, singing abilities, and emotions) are associated with vocal emotional processing. Finally, we examined whether any association between music training and vocal emotions is mediated by perceptual skills (music training → perceptual skills → vocal emotion recognition). Complete mediation would imply that the association depends primarily on relatively low-level listening skills, which music training may enhance. By contrast, partial or no mediation would imply that the association between music and vocal emotions is also driven by nonperceptual processes, possibly at higher-order cognitive or social levels (e.g., emotional and social components of music activities).

Method

Ethical approval for the study protocol was obtained from the Departmental Ethics Committee, Faculty of Psychology and Education Sciences, University of Porto (reference 3–1/2017). Written informed consent was collected from all participants, who were either paid or given partial course credit.

Participants

A total of 172 participants were recruited from research participant pools or in response to advertisements on campus or on social media. Three were excluded for not completing the Gold-MSI, which resulted in a final sample of 169 (116 female). Participants were 23.49 years of age on average (*SD* = 8.27, range = 18–72). According to self-reports, all had normal hearing and no history of neurological or psychiatric disorders, and all were native speakers of European Portuguese. Formal music training varied widely, as illustrated in Figure 1. The mode was no training (n = 69), but 100 had some training, ranging from 0.5 to 10 or more years. Duration of music training was not associated with age (r = −.01, p = .87, BF_{10} = 0.10) or sex (r = −.12, p = .11, BF_{10} = 0.35), but it had a very weak association with education (r = .18, p = .02, BF_{10} = 1.36). In the statistical analyses, we considered duration of music training in two ways: as a continuous 7-point variable (as measured by the Gold-MSI, see below), and with group comparisons between highly trained participants and those with no training, which is the norm in this line of research (for a review, Schellenberg, 2019). Participants with six or more years of instrumental training were considered to be highly trained (n = 30), as in previous research (Zhang, Susino, McPherson, & Schubert, 2018).

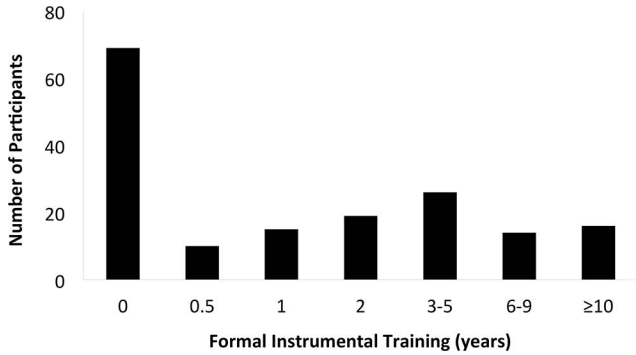

Figure 1. Distribution of formal instrumental training across participants.

Power analysis (with G*Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009) indicated that for our main analyses, a sample of at least 134 participants was required to be 95% certain of detecting partial associations of r = .30 or larger between each predictor variable and emotion recognition accuracy, using multiple-regression models that included three predictors (music training, music perception abilities, and digit span).

Materials

Self-reported musical abilities. The Gold-MSI includes 38 items that cover a wide variety of music skills, expertise, and behaviors (Müllensiefen et al., 2014). It is suited for measuring individual differences among performing musicians as well as among members of the general population who vary in musical skills and interest in music. Scale items are grouped into five subscales, each of them corresponding to a different facet of musicality: active engagement (nine items; e.g., *I spend a lot of my free time doing music-related activities*), perceptual abilities (nine items, e.g., *I can tell when people sing or play out of tune*), music training (seven items, e.g., *I have had formal training in music theory for __ years*), singing abilities (seven items, e.g., *I am able to hit the right notes when I sing along with a recording*), and emotion (six items, e.g., *I am able to talk about the emotions that a piece of music evokes for me*). For the first 31 items, participants indicate their level of agreement with each statement using a seven-point Likert scale (from 1 = *completely disagree* to 7 = *completely agree*). For the remaining items, participants use ordinal scales with seven response alternatives (e.g., *I can play [number from 0 to '6 or more'] instruments*). Thus, for each participant, each original item is scored with an integer that ranges from 1 to 7.

The Gold-MSI and its translation have good psychometric properties (Lima et al., 2018; Müllensiefen et al., 2014;). Construct validity has been documented with associations between index scores and performance-based music perception tasks (i.e., beat

alignment and melody memory, Müllensiefen et al., 2014; discrimination of pitch and duration, Dawson, Aalto, Šimko, Vainio, & Tervaniemi, 2017).

Performance-based auditory and musical abilities. Four tasks were used to measure musical beat perception, melodic memory, pitch discrimination, and duration discrimination. The musical beat and melodic memory tasks were optimized versions of the ones used by Müllensiefen et al. (2014). For the *beat alignment* test, stimuli were 17 short excerpts of music (10–16 s), which included a beep track similar to a metronome. The beep track coincided with the implied beat of the music on four trials. On the other 13, the beep was phase shifted by 10 or 17.5%, or changed in tempo by 2%. On each trial, participants indicated whether the beat track was *on* or *off* the beat, as in the Beat Alignment Test (Iversen & Patel, 2008). The order of trials was randomized across participants.

For the *melodic memory* task (Harrison, Musil, & Müllensiefen, 2016), participants listened to 13 pairs of short tunes (10–17 notes) and determined whether each pair was the *same* or *different*. The second tune was always transposed by 1 or 7 semitones. Thus, the task required listeners to determine whether both melodies had the same structure of consecutive musical intervals. Five pairs had a *different* structure, in which 1–3 notes were changed (as in Bartlett & Dowling, 1980; Cuddy & Lyons, 1981) to alter the contour and intervals, or maintain the contour but change the intervals. Both the musical beat and melodic memory tasks were implemented in PsychoPy Experiment Builder v1.85.4 (http://www.psychopy.org/) by Estela Puig-Waldmüller and Bruno Gringras (University of Vienna), with Portuguese instructions. Each task took approximately 7 min to complete.

For the pitch and duration tasks, discrimination thresholds were obtained from a two-down-one-up adaptive staircase procedure, which tracked good but not perfect performance (70.7% correct) on the psychometric function (Soranzo & Grassi, 2014). For pitch discrimination, participants were presented with trials consisting of three consecutive 250 ms pure tones: two of them had the same frequency (always 1000 Hz) and one was higher. The difference was 100 Hz at the beginning, but subsequently varied from 2 to 256 Hz based on the listener's performance. Correct identification of the higher tone led to progressively smaller pitch differences until participants stopped responding correctly, whereas incorrect responses led to progressively larger differences until they responded correctly. For duration discrimination, listeners heard three pure tones on each trial, and judged which was the longest. Two of the tones were always 250 ms and one was longer by 100 ms at the beginning, but subsequently varied between 8 and 256 ms. For both tasks, the procedure ended after 12 reversals (i.e., changes in the direction of the stimulus difference). Thresholds were calculated by averaging the values of the last eight reversals. Lower values indicated better performance. The PSYCHOACOUSTICS toolbox (Soranzo & Grassi, 2014) running on MATLAB (R2016b, Version 9.1.0) was used for both tasks, and each of them took approximately 5 min to complete, although duration varied depending on performance.

Emotion recognition. Participants completed three emotion recognition tasks that were identical except for the stimuli. Two were auditory, with emotions conveyed through nonverbal vocalizations or speech prosody. The third was visual (facial expressions). Each task had 84 trials, with 12 different stimuli representing each of seven emotions (anger, disgust, fear, happiness, pleasure/surprise, sadness, and neutral). The stimuli were taken from previously validated corpora (speech prosody, Castro & Lima, 2010; nonverbal vocalizations, Lima, Castro, & Scott, 2013; facial expressions, Karolinska Directed Emotional Faces database, Goeleven, De Raedt, Leyman, & Verschuere, 2008; Lundqvist, Flykt, & Öhman, 1998) and have been used frequently (e.g., Eisenbarth & Alpers, 2011; Lima et al., 2013, 2016; Lima & Castro, 2011; Strachan, Sebanz, & Knoblich, 2019).

Speech prosody stimuli were short sentences (M = 1472 ms, SD = 247) with emotionally neutral semantic content (e.g., "O quadro está na parede", *The painting is on the wall*), produced by two female speakers to communicate emotions with prosodic cues alone (i.e., variations in pitch, loudness, timing, and voice quality). Nonverbal vocalizations consisted of brief vocal sounds (M = 1013 ms, SD = 286) without verbal content, such as laughs, screams, or sobs, as produced by two female and two male speakers. Finally, facial expressions consisted of color photographs of male and female actors with no beards, moustaches, earrings, eyeglasses, or visible make-up. Each photograph was presented for 2 s. The three tasks were similarly difficult. Based on validation data from the different corpora, average recognition accuracy was 75.6% for speech prosody, 80.7% for nonverbal vocalizations, and 79.4% for facial expressions.

Participants made an eight-alternative forced-choice judgment for each stimulus in each task, selecting the emotion that was being expressed from a list that included *neutrality*, *anger*, *disgust*, *fear*, *happiness*, *pleasure/surprise*, *sadness*, and *none of the above*. Each of the three tasks started with four practice trials. The 84 experimental trials that followed were randomized separately for each participant. Each stimulus was presented once and no feedback was provided. The tasks were implemented in E-Prime 2.0 Professional (Version 2.0.10.356), and each took approximately 10 min to complete.

General cognitive ability. To index domain-general cognitive abilities, participants completed the forward and backward portions of the Digit Span subtest from the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 2003). A total score was the sum of the forward and backward raw scores.

Procedure

Participants were tested individually in a quiet room. They completed a background questionnaire that asked for demographic information, and then the remaining questionnaires, the experimental tasks, and the digit span test. The order of the tasks was randomized across participants, and the testing session lasted about 1.5 hr. Short breaks were allowed between tasks. The auditory stimuli were presented via high-quality headphones (Sennheiser HD 280 Professional), with the volume adjusted to a comfortable level for each participant.

The same participants also completed a task that required them to compare the emotional features of pairs of musical excerpts (MacGregor & Müllensiefen, 2019), and a series of questionnaires that indexed emotion- and health-related variables. These results will be reported in a separate publication.

Some data were missing for some tasks: beat perception (n = 11), melodic memory (n = 17), pitch discrimination (n = 3), duration discrimination (n = 1), emotional prosody (n = 4),

vocalizations ($n = 2$), faces ($n = 5$), and digit span ($n = 1$). Thus, the sample size varied slightly in the statistical analyses depending on which variables were involved.

Data Preparation and Analysis

Because we had four performance-based music perception tasks (musical beat perception, melodic memory, pitch discrimination, and duration discrimination), we asked whether an aggregate variable could be formed and used as an index of musical ability to reduce collinearity and the contribution of measure-specific error variance. A principal component analysis (varimax rotation) revealed that a two-factor solution accounted for 73% of the variance in the original data. Three of the tasks loaded highly on the first component (beat perception, pitch discrimination, and duration discrimination, rs $= -.76$, .79, and .81, respectively), whereas melodic memory was almost perfectly correlated with the second component ($r = .98$). In the analyses that follow, we used the original *melodic memory* accuracy scores, and an aggregate *music perception* variable that represented the principal component extracted from the other three variables, which was almost perfectly correlated with the first component from the original analysis ($r = .98$). Lower scores on this measure indicate better performance.

Accuracy rates for emotion recognition tasks were arcsine square-root transformed and corrected for possible response biases using unbiased hit rates, or Hu (Wagner, 1993; for a discussion of biases in forced-choice tasks, e.g., Isaacowitz et al., 2007). Hu values represent the joint probability that a given emotion will be correctly recognized (given that it is presented), and that a given response category will be correctly used (given that it is used at all), such that they vary between 0 and 1. Hu = 0 when no stimulus from a given emotion is correctly recognized, and Hu = 1 only when all the stimuli from a given emotion (e.g., happy prosody) are correctly recognized, and the corresponding response category (e.g., happiness) is always correctly used (i.e., when there are no false alarms). Most analyses were conducted using average scores for each task because we had no predictions regarding specific emotions. For comparisons between trained and untrained participants, however, we also tested for the possibility of emotion-specific effects, to ensure that associations with training were not driven by a single emotion or a small subset of emotions, as they have been in previous research (Pinheiro et al., 2015; Thompson et al., 2004).

The data were statistically evaluated based on standard frequentist *and* Bayesian approaches (e.g., Jarosz & Wiley, 2014). In each analysis, a Bayes Factor (BF_{10}) statistic was estimated, which considers the likelihood of the observed data given the alternative and null hypotheses. These analyses were conducted on JASP Version 0.9.2 (JASP Team, 2018), using default priors (Rouder, Morey, Speckman, & Province, 2012; Wagenmakers, Love, et al., 2018; Wagenmakers, Marsman, et al., 2018; Wagenmakers, Verhagen, & Ly, 2016). BF_{10} values were interpreted following Jeffreys' guidelines (Jarosz & Wiley, 2014; Jeffreys, 1961), with values between 1 and 3 corresponding to weak or anecdotal evidence for the alternative hypothesis, values between 3 and 10 corresponding to substantial evidence, and values between 10 and 30, 30 and 100, and over 100 corresponding to strong, very strong, and decisive evidence, respectively. A $BF_{10} < 1$ provided evidence in favor of the null hypothesis ($BF_{10} < 0.33$ indicated substantial evidence, $BF_{10} < 0.10$ indicated strong evidence).

Results

Formal Music Training

Table 1 shows summary statistics for the full sample, for highly trained individuals only ($n = 30$), and for participants with no training ($n = 69$). Table 2 provides correlations between duration of music training and the other variables using the full sample. As in previous studies, music training was associated robustly with enhanced musical abilities on both self-report and performance-

Table 1
Descriptive Statistics for the Full Sample, and Separately for Highly Musically Trained and Untrained Participants

Task	Full sample ($N = 169$) *M* (*SD*)	Untrained ($n = 69$) *M* (*SD*)	Highly Trained ($n = 30$) *M* (*SD*)	*p*-value (BF_{10})
Digit span (WAIS-III; total)	15.11 (3.67)	14.58 (4.20)	16.69 (3.13)	.02 (2.94)
Gold-MSI (Likert scale, 1–7)				
Music training	3.14 (1.68)	1.72 (0.71)	5.66 (0.77)	<.001 (>100)
Perceptual abilities	5.23 (0.94)	4.88 (0.95)	6.06 (0.63)	<.001 (>100)
Active engagement	4.19 (1.10)	3.87 (1.10)	4.86 (1.05)	<.001 (>100)
Singing abilities	4.23 (1.13)	3.85 (1.04)	5.14 (0.88)	<.001 (>100)
Emotions	5.71 (0.77)	5.57 (0.86)	5.98 (0.60)	.02 (2.38)
Music perception tasks				
Aggregate music perception	0.00 (1.00)	0.31 (0.97)	−0.79 (0.75)	<.001 (>100)
Melodic memory (% correct)	63.51 (14.23)	59.10 (14.82)	73.08 (15.17)	<.001 (>100)
Emotion recognition (average Hu scores)				
Prosody	.63 (.17)	.61 (.17)	.69 (.17)	.02 (2.70)
Vocalizations	.78 (.11)	.76 (.13)	.81 (.10)	.02 (2.65)
Faces	.70 (.10)	.69 (.12)	.70 (.09)	.93 (0.15)

Note. WAIS-III = Wechsler Adult Intelligence Scale-Third Edition; Gold-MSI = Goldsmiths Musical Sophistication Index. For the Gold-MSI and music perception tasks, *p* values correspond to the statistic of independent samples *t* tests (two-tailed). For the emotion recognition tasks, *p* values correspond to the main effect of group in mixed-design analysis of variances (ANOVAs), including music training as between-subject factor and emotion as repeated-measures factor.

Table 2
Pairwise Correlations Between Duration of Music Training and Other Variables

Variable	*r*	*p*-value	BF_{10}
Digit span (total)	.18	.02	1.60
Gold-MSI: Music training	.89	<.001	>100
Gold-MSI: Perceptual abilities	.42	<.001	>100
Gold-MSI: Active engagement	.32	<.001	>100
Gold-MSI: Singing abilities	.44	<.001	>100
Gold-MSI: Emotions	.18	.02	1.33
Aggregate music perception	−.39	<.001	>100
Melodic memory	.34	<.001	>100
Emotion recognition: Prosody (average)	.21	.01	3.71
Emotion recognition: Vocalizations (average)	.21	.01	4.25
Emotion recognition: Faces (average)	.10	.18	0.24

Note. Gold-MSI = Goldsmiths Musical Sophistication Index.

based tasks. Associations with general cognitive abilities were evident but weak.

As expected, duration of music training was correlated positively with average emotion recognition scores across the full sample, both for speech prosody and for nonverbal vocalizations (see Table 2). The effect was small ($r = .21$ in both cases), but Bayesian analyses indicated that the evidence was substantial. By contrast, for facial expressions, there was substantial evidence for a *null* effect. When digit span was held constant, associations between duration of training and emotion recognition remained significant for nonverbal vocalizations (partial $r = .22$, $p = .01$, $BF_{10} = 9.45$), and at marginal levels for speech prosody (partial $r = .15$, $p = .06$, $BF_{10} = 1.03$).

We then conducted group comparisons between highly trained participants and those without any training. Mixed-design analyses of variance (ANOVAs) were conducted for each task, with the different emotions as a repeated measure, and music training as a between-subjects factor. Greenhouse-Geisser corrections were applied when necessary (Mauchly's sphericity test). For speech prosody, we found a significant advantage for highly trained participants, $F(1, 95) = 5.80$, $p = .02$, $\eta^2 = .06$, although Bayesian statistics suggested that the evidence was weak, $BF_{10} = 2.70$. A main effect of emotion category confirmed that some emotions were more difficult to recognize than others, $F(4.98, 472.73) = 26.42$, $p < .001$, $\eta_p^2 = .22$, $BF_{10} > 100$, but there was no interaction between music training and emotion, $p = .25$, $BF_{10} = .05$. For nonverbal vocalizations, an advantage for trained participants was again observed, $F(1, 96) = 6.11$, $p = .02$, $\eta^2 = .06$, $BF_{10} = 2.65$. The main effect of emotion was significant, $F(4.30, 412.82) = 10.26$, $p = .001$, $\eta_p^2 = .10$, $BF_{10} > 100$, as was the interaction between music training and emotion, $F(4.30, 412.82) = 4.11$, $p = .002$, $\eta_p^2 = .04$, $BF_{10} = 37.32$. Follow-up analyses showed that trained participants were numerically better than their untrained counterparts for all emotions except anger (average Hu scores = .76 and .78 for trained and untrained participants, respectively, $p = .37$). Finally, for facial expressions, there was substantial evidence for a *null* effect of music training, $p = .93$, $BF_{10} = .15$, a main effect of emotion, $F(4.57, 439.29) = 59.83$, $p < .001$, $\eta_p^2 = .39$, $BF_{10} > 100$, but no interaction between training and emotion, $p = .41$, $BF_{10} = .03$.

When digit span was held constant, the advantage for highly trained participants remained significant for nonverbal vocalizations, $F(1, 94) = 6.15$, $p = .01$, $\eta_p^2 = .06$, $BF_{10} = 3.06$, as did the main effect of emotion, $p = .17$, $BF_{10} > 100$, and the interaction between training and emotion, $F(4.35, 409.23) = 2.76$, $p = .02$, $\eta_p^2 = .03$, $BF_{10} = 35.04$. For speech prosody, however, the main effect of emotion remained evident, $F(4.92, 457.55) = 4.46$, $p < .001$, $\eta_p^2 = .05$, $BF_{10} > 100$, but the advantage for trained participants disappeared, $F(1, 93) = 2.31$, $p = .13$, $BF_{10} = 0.60$, and there was no interaction between training and emotion, $p = .23$, $BF_{10} = .08$.

In short, we found evidence for a positive association between music training and the recognition of emotion in voices but not faces. The effect was small, however, and in the case of prosody it was partly related to individual differences in digit span.

Self-Reported Musical Abilities

We then tested for associations between emotion recognition and facets of musical abilities other than music training, as assessed by the subscales from the Gold-MSI. Zero-order correlations are provided in the upper part of Table 3. As predicted, we found decisive evidence that higher music perception abilities correlated with higher emotion recognition accuracy. These associations were observed for prosody and nonverbal vocalizations, but not for facial expressions. Exploratory analyses also revealed

Table 3
Correlations Between Emotion Recognition and Musical Abilities as Measured With Gold-MSI Subscales and Objective Music Perception Tasks

	Prosody		Vocalizations		Faces	
Task	*r*	BF_{10}	*r*	BF_{10}	*r*	BF_{10}
Gold-MSI						
Perceptual abilities	.32**	>100	.34**	>100	.15	0.63
Active engagement	.20	2.73	.21	4.25	.11	0.24
Singing abilities	.30*	>100	.22	6.11	.16	0.72
Emotions	.14	0.48	.18	1.27	.18	1.33
Music perception tasks						
Aggregate music perception	−.39**	>100	−.42**	>100	−.18	1.03
Melodic memory	.12	0.28	.12	0.31	.22	4.01

Note. Gold-MSI = Goldsmiths Musical Sophistication Index.
* $p < .05$. ** $p < .001$ (Holm Bonferroni-corrected).

an unpredicted association between singing abilities and emotion recognition performance, but only for speech prosody.

An important question was whether the association between music perception and vocal emotion recognition would remain evident when music training and general cognitive abilities were held constant. Using multiple regression, we modeled average accuracy on the speech prosody task as a function of music perception abilities, duration of music training, and digit span. The model explained 14.6% of the variance, $R = .38$, $F(3, 162) = 9.23$, $p < .001$, $BF_{10} > 100$. Independent contributions to the model were evident for music perception, partial $r = .18$, $p = .02$, $BF_{10} = 2.92$, and digit span, partial $r = .23$, $p = .003$, $BF_{10} = 12.54$, but not for music training, $p = .38$, $BF_{10} = 0.33$. A similar regression analysis was conducted for nonverbal vocalizations. The three-predictor model explained 14.3% of the variance, $R = .38$, $F(3, 163) = 9.06$, $p < .001$, $BF_{10} > 100$. Music perception abilities made a decisive independent contribution to the model, partial $r = .32$, $p < .001$, $BF_{10} > 100$, digit span contributed anecdotally, partial $r = -.16$, $p = .04$, $BF_{10} = 1.86$, but music training was irrelevant, $p = .16$; $BF_{10} = 1.76$. Associations between self-reported music perception abilities and vocal emotion recognition are illustrated in the upper part of Figure 2.

We also confirmed that self-reported music perception abilities predicted unique variance in vocal emotion recognition even when age, sex, and education were added to the regression models (in addition to music training and digit span), $ps \leq .02$, all $BF_{10} > 3.17$.

Performance-Based Musical Abilities

In the next set of analyses, we asked whether similar findings could be observed for objective measures of music perception abilities. As shown in the lower part of Table 3, no associations were found for melodic memory, but we found decisive evidence for a positive association in the case of the aggregate measure. Participants with higher music perception abilities also had improved emotion recognition for prosody and nonverbal vocalizations, but not for facial expressions.

Multiple-regression models showed that these associations remained evident when music training and digit span were held constant. For speech prosody, a model with three predictor variables (aggregate measure of music perception, duration of music training, and digit span) accounted for 18.8% of the variance, $R = .43$, $F(3, 148) = 11.44$, $p < .001$, $BF_{10} > 100$. Independent contributions were made by the aggregate measure of music perception, partial $r = -.31$, $p < .001$, $BF_{10} > 100$, and digit span, partial $r = .21$, $p = .01$, $BF_{10} = 5.37$, but there was no contribution of music training, $p = .93$; $BF_{10} = 0.22$. A similar model for nonverbal vocalizations accounted for 20.5% of the variance $R = .45$, $F(3, 149) = 12.79$, $p < .001$, $BF_{10} > 100$. Independent contributions were again evident for the aggregate measure of music perception, partial $r = -.40$, $p < .001$, $BF_{10} > 100$, and digit span, partial $r = -.18$ $p = .03$, $BF_{10} = 2.28$, but not for music training, $p = .31$, $BF_{10} = 0.34$. Associations between objective music perception abilities and vocal emotion recognition are illustrated in the lower part of Figure 2.

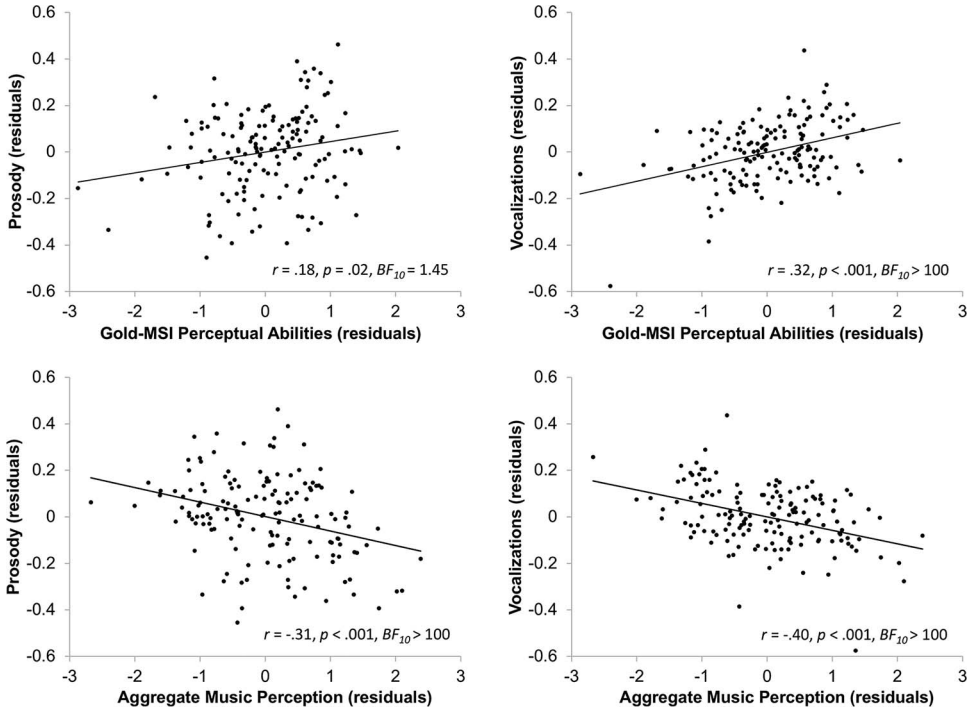

Figure 2. Partial regression plots illustrating the relationship between music perception abilities (Goldsmiths Musical Sophistication Index [Gold-MSI] Perceptual Abilities and Aggregate Music Perception) and vocal emotion recognition (prosody and vocalizations), after removing the effects of music training and digit span. Lower scores in Aggregate Music Perception indicate better performance.

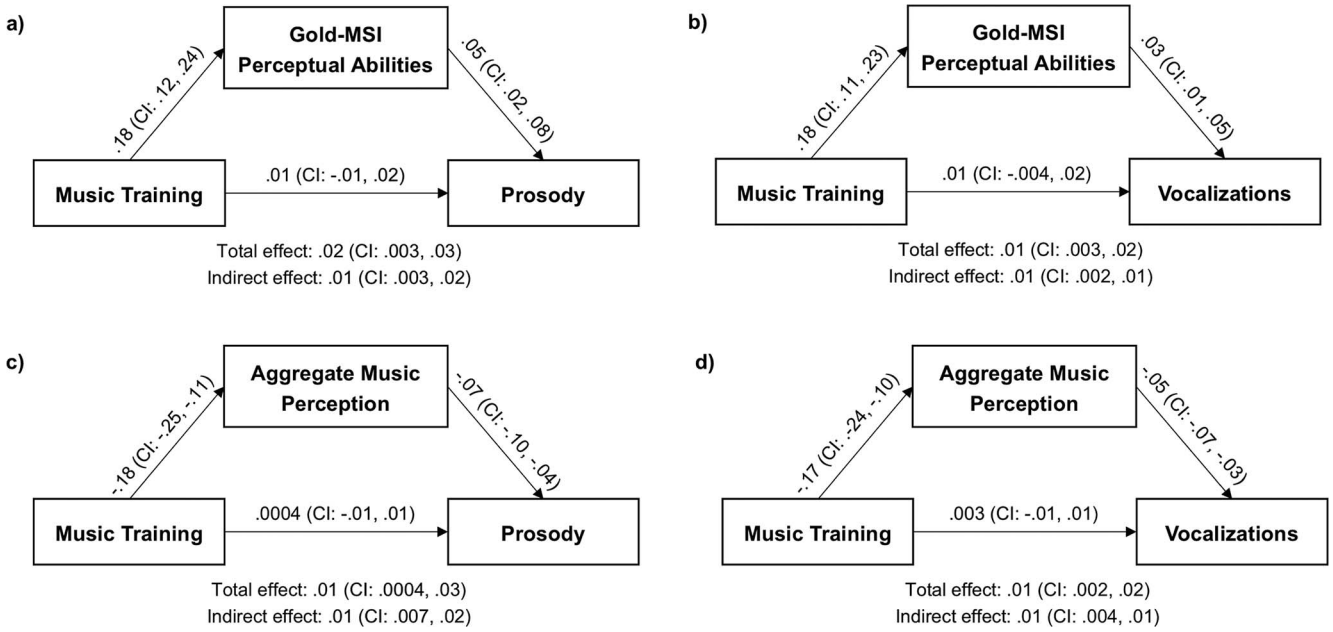

Figure 3. Models depicting the mediation effect of music perception abilities (Goldsmiths Musical Sophistication Index [Gold-MSI] Perceptual Abilities and Aggregate Music Perception) on the association between music training and vocal emotion recognition (prosody and vocalizations). Inference was based on percentile bootstrap 95% confidence intervals (CIs) with 20.000 samples. Lower scores in Aggregate Music Perception indicate better performance.

The aggregate measure of music perception abilities predicted unique variance in vocal emotion recognition even when age, sex, and education were also included in the regression models, $ps < .001$, all $BF_{10} > 100$.

Nonmusicians With Good Musical Abilities Versus Highly Trained Participants

The previous analyses established that individuals with better music perception abilities are also better at recognizing vocal emotions, regardless of music training. An interesting question is whether untrained participants with good musical abilities show emotion recognition performance comparable to that of trained musicians. To address this question, we divided untrained participants into high and low musical abilities groups, based on median-splits of the music perception measures. Separate analyses were conducted for self-reported music perception scores and the aggregate measure of music perception. We then compared those with high musical abilities with trained musicians.[1] For speech prosody, there was no advantage for trained participants: self-reports, $t(63) = -1.48$, $p = .14$, $BF_{10} = 0.64$; performance-based skills, $t(58) = -1.98$, $p = .06$, $BF_{10} = 1.30$. Similarly, for nonverbal vocalizations, highly trained participants did not differ from untrained ones with good musical abilities: self-reports, $t(63) = -1.54$, $p = .69$, $BF_{10} = 0.65$; performance-based skills, $t(58) = -1.39$, $p = .17$, $BF_{10} = 0.59$. In short, musician-like enhancements in vocal emotion recognition were evident in participants without any formal music training, provided they had good musical abilities.

Mediation Analyses

A final analysis determined whether the association between music training and emotion recognition was mediated by music perception skills, which are enhanced in trained individuals. These analyses were conducted using the PROCESS macro for SPSS (Version 3.3; Hayes, 2017), with statistical inferences based on percentile bootstrap 95% confidence intervals (CIs) with 20,000 samples. Total, direct, and indirect (mediated) effects were estimated, and were considered significant when the CIs did not include 0.

The mediation models are depicted in Figure 3. For speech prosody, the indirect effect of music training on emotion recognition scores—through self-reported music perception skills—was significant. The direct effect was not, however, indicating that there was no association between training and emotion recognition performance when music perception skills were held constant. Identical results emerged when the objective music perception measure (aggregate measure) was substituted for the self-reported one, as well as in similar analyses for nonverbal vocalizations. In short, duration of music training was positively associated with enhanced emotion recognition because trained individuals had enhanced music perception skills.

[1] Nonmusicians with good musical abilities and trained musicians were generally similar in digit span and sample size. Nonmusicians with good musical abilities: based on self-reports, $n = 36$, digit span $M = 15.19$; based on performance-based skills, $n = 32$, digit span $M = 15.09$. Trained musicians: based on self-reports, $n = 30$, digit span $M = 16.69$; based on performance-based skills, $n = 29$, $M = 16.75$.

Discussion

The present study examined the association between musical expertise and the ability to recognize emotions in vocal and facial expressions. We examined associations with formal music training and, crucially, we investigated whether, in the absence of training, having good musical abilities related to enhancements in emotion recognition similar to the ones seen in musicians. The analyses had four main findings. First, music training was associated with better emotion recognition in speech prosody and nonverbal vocalizations. The advantage was small, though, and restricted to the auditory domain (i.e., not observed for facial emotion recognition). Second, we found a robust association between music perception skills and enhanced vocal emotion recognition, which remained significant even when music training and general cognitive abilities were held constant. Untrained participants with good musical abilities showed vocal emotion recognition performance comparable to that of trained musicians. Third, self-reported singing abilities were associated positively with emotional prosody recognition. Fourth, mediation analyses showed that the effect of music training on vocal emotion recognition was fully mediated by music perception skills.

In previous studies, an advantage for musicians in emotional prosody recognition has been reported in some instances (Lima & Castro, 2011; Thompson et al., 2004), but not in others (Mualem & Lavidor, 2015; Park et al., 2015; Trimmer & Cuddy, 2008). The present study was the first to examine whether musicianship predicts the recognition of emotions in other types of vocal expressions. Our results corroborated the association with speech prosody and extended it to nonverbal vocalizations. The advantage for vocalizations was consistent with previous evidence showing a more efficient subcortical encoding of crying sounds in musicians compared with nonmusicians (Strait et al., 2009). In short, associations with music training can be seen for vocal emotional stimuli with or without linguistic information, which suggests that it might stem from a general benefit in decoding vocal emotional cues, because emotional speech and nonverbal vocalizations differ in their production and perceptual mechanisms (Scott et al., 2010). For instance, electrophysiological evidence indicates that vocalizations and emotional prosody are differentiated rapidly, with preferential processing of vocalizations at early stages of auditory processing and allocation of attention (N1-P2 components; Pell et al., 2015). The failure of some previous studies to replicate the musicians' advantage in emotion recognition could be because the association is small and relatively weak, as suggested by our Bayesian analyses. In other words, a relatively large sample of highly trained participants might be required for such an association to emerge. In previous studies with null findings, the samples included less than 15 musicians (Park et al., 2015; Pinheiro et al., 2015), or participants with only a modicum of training.

We also found that, in the case of speech prosody (but not in the case of vocalizations), the association with music training became nonsignificant after accounting for individual differences in digit span. This finding implies that it is a consequence of domain-general cognitive abilities. In an earlier study, however, Lima and Castro (2011) documented that associations between musical expertise and prosody were independent of general abilities. This discrepancy might stem from differences in samples, or from the particular way domain-general abilities are measured. Our digit span task measured auditory-perceptual processes, which could, arguably, be a consequence of the training itself rather than a proper confounding variable. By contrast, Lima and Castro (2011) had several domain-general cognitive tasks, including purely nonverbal ones such as Raven's Advanced Progressive Matrices. The precise role of different domain-general processes could be addressed in future studies.

A novel but null finding of the present study was that music training had no association with emotion recognition in the visual domain. Thus, musicians' advantage in emotion recognition may be restricted to the auditory domain. Our decision to include a facial emotion recognition task was motivated by evidence of domain-general socioemotional processing difficulties in congenital amusia (Lima et al., 2016). There is also evidence that musicians show stronger responses to emotional prosody in brain regions involved in modality-independent inferences about mental states, including the medial prefrontal and anterior cingulate cortices (Park et al., 2015). Moreover, socioemotional stimuli in everyday life are typically multimodal, such that individual differences in basic auditory skills could have cascading effects that extend to higher-order aspects of socioemotional cognition. Perhaps the enhancements associated with good musical abilities are relatively small and, thus, incapable of being observable across domains. In the case of congenital amusia, affected individuals have severe pitch deficits that are likely to influence early stages of socioemotional development. Our null results for music training and visual stimuli are consistent with recent meta-analyses that raise doubts about far transfer in general, and as a consequence of music training in particular (Sala & Gobet, 2017a, 2017b). Indeed, in some instances, there is no association between music training and performance on nonmusical *auditory* tasks such as perceiving speech in noise (Boebinger et al., 2015; Madsen et al., 2019).

By assessing basic auditory and music perception skills in addition to music training, we were able to find robust evidence that being a musician is not a necessary condition for music-related advantages in vocal emotion recognition. Converging data from self-report and performance-based measures indicated, moreover, that auditory and musical skills are broadly associated with enhanced emotional processing of speech prosody and nonverbal vocalizations, even after accounting for training. These findings suggest that neurocognitive pathways for music and vocal emotions overlap, and that such overlap stems from aspects of musical expertise other than formal training. Crucially, they establish a role for factors other than formal training in associations between musical and nonmusical abilities, specifically listening abilities that are driven by innate predispositions and/or informal engagement with music.

Our findings also align well with recent evidence of associations between music and speech and language processing. For example, Swaminathan and Schellenberg (2017, 2019) found that for adults and children, rhythm perception abilities were associated positively with phoneme discrimination in a foreign language, even after controlling for music training and domain-general cognitive abilities. In other words, music perception skills (i.e., rhythm in these instances) were a better predictor of speech perception than music training.

Mankel and Bidelman (2018) examined neuroelectric brain responses to clear and noise-degraded speech sounds in musically untrained participants. Participants with higher music perception

skills had more tightly linked frequency-following responses to speech, and more faithful representations of speech in noise. Although the authors proposed that music training provides an additional boost, on top of preexisting skills, to the neural processing of speech, they had no clear evidence for this assertion, and we did not find evidence for such a boost in the current study. Rather, our highly adept nonmusicians showed vocal emotion recognition performance that was on par with highly trained musicians. Perhaps the level of analysis plays a role. Mankel and Bidelman (2018) emphasized neural measures, whereas our evidence was behavioral. In fact, Mankel and Bidelman found different results for neural and behavioral measures: although neural measures were sensitive to fine differences in musical skills, a behavioral measure was not. In short, advantages related to musical abilities (in the absence of training) are not discernible from those attributed to training at a behavioral level.

Although we documented an important role for predispositions and informal experience in the association between musical abilities and vocal emotion recognition, we do not doubt that music training can produce experience-dependent effects. Carefully designed longitudinal studies with random assignment provide evidence for such effects across a range of tasks, at behavioral and neural levels (Bangert & Altenmüller, 2003; Chobert, François, Velay, & Besson, 2014; François, Chobert, Besson, & Schön, 2013; Frey et al., 2019; Moreno et al., 2009, 2011; Schellenberg, 2004), yet more research is crucial to clarify their robustness and scope. Moreover, group differences that are evident in cross-sectional studies have often been causally attributed to training (Schellenberg, 2019), yet the current findings confirm that similar advantages can be seen in individuals without any training. By equating musical expertise with music training, neuroscientists and psychologists are exhibiting a limited perspective of the genetic and environmental factors that shape musical abilities, and of the richness and diversity of musical behaviors and experience. A complete understanding of musicality and its role in cognition requires a complex and multifaceted exploration of musical skills.

In exploratory analyses, we found that self-reports of singing abilities were associated positively with recognizing emotions in speech prosody. This finding implies that other facets of musical expertise are involved in associations with vocal emotions. It is also consistent with well-documented behavioral and neural links between production, imagery, and perceptual mechanisms that are involved in voice processing (Correia et al., 2019; Lima, Krishnan, & Scott, 2016; Lima et al., 2015; McGettigan et al., 2015; Pfordresher & Halpern, 2013; Warren et al., 2006). On the one hand, vocal production (such as singing) involves not only implementing movements, but also planning and anticipating outcomes (that might rely on imagery), and using auditory feedback (perceptual processes) to detect and correct errors. On the other hand, listening to sounds, namely vocal emotional sounds, recruits auditory areas in the temporal lobes as well as areas involved in motor planning and control. The involvement of the motor system is also correlated positively with enhanced vocal emotional processing (Correia et al., 2019; McGettigan et al., 2015), which suggests that more efficient activation of sound-related motor representations optimizes perceptual processes (for review see Lima et al., 2016). Therefore, it is plausible that such tight production-perception links in voice processing account for the positive association between singing and speech prosody perception, a prediction that could be tested systematically in future studies. For example, singing abilities could be assessed not only via self-report but also with performance-based tasks (e.g., Pfordresher & Halpern, 2013).

Previous studies provided suggestive evidence that the primary locus for transfer effects from music training to vocal emotions is at a basic auditory-perceptual level of processing. Music training could lead to more efficient auditory-perceptual processing (Herholz & Zatorre, 2012; Kraus & Chandrasekaran, 2010), which in turn could facilitate vocal emotion recognition, because sensory processing is central to vocal communication (Schirmer & Kotz, 2006). Our mediation analysis provided results consistent with this view, in the sense that no other component of musical expertise played an important role. Indeed, the effect of training on vocal emotion recognition was accounted for entirely by advantages in music perception skills. This finding can be interpreted within the framework of the OPERA hypothesis, which describes transfer effects from music training to the encoding of speech (Patel, 2011, 2014). Because the brain networks that process acoustic features are used in both music and speech, musicians may exhibit speech processing benefits because their training includes several features that could modulate these networks (precision, emotion, repetition, and attention). It is noteworthy that the association between auditory-perceptual skills and emotion processing did not extend to the visual domain, which implies that the overlap between musical skill and emotion recognition does not extend to higher-order levels of processing. Future studies using techniques such as electroencephalogram (EEG) or functional magnetic resonance imaging (fMRI) could be useful to address these questions more directly, because they can tell us *when* cross-domain interactions occur (early vs. later stages of processing), and if they occur primarily in auditory areas or, rather, extend to regions involved in supramodal socioemotional processing.

To conclude, the present study represents the first demonstration that better music perception skills are associated with enhancements in vocal emotion recognition, even in the absence of any formal music training. Untrained individuals who are naturally musical can be as good as highly trained musicians at recognizing emotions in speech prosody and nonverbal vocalizations. Our findings do not rule out the possibility that music training induces experience-dependent effects, but they affirm an important role of preexisting factors in associations between music and nonmusical domains that have been neglected in the literature. Collectively, the results reported here emphasize the need to interpret cross-sectional music training effects with caution. They also confirm that there are multiple facets to musical expertise beyond formal training, which future research could examine in greater detail.

References

Bangert, M., & Altenmüller, E. O. (2003). Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neuroscience, 4,* 26. http://dx.doi.org/10.1186/1471-2202-4-26

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 501–515. http://dx.doi.org/10.1037/0096-1523.6.3.501

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100,* 100–130. http://dx.doi.org/10.1016/j.cognition.2005.11.007

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America, 137,* 378–387. http://dx.doi.org/10.1121/1.4904537

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews, 8,* 383–403. http://dx.doi.org/10.1016/j.plrev.2011.10.002

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods, 42,* 74–81. http://dx.doi.org/10.3758/BRM.42.1.74

Chobert, J., François, C., Velay, J. L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex, 24,* 956–967. http://dx.doi.org/10.1093/cercor/bhs377

Clark, C. N., Downey, L. E., & Warren, J. D. (2015). Brain disorders and the biological role of music. *Social Cognitive and Affective Neuroscience, 10,* 444–452. http://dx.doi.org/10.1093/scan/nsu079

Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in Neuroscience, 11,* 479. http://dx.doi.org/10.3389/fnins.2017.00479

Correia, A. I., Branco, P., Martins, M., Reis, A. M., Martins, N., Castro, S. L., & Lima, C. F. (2019). Resting-state connectivity reveals a role for sensorimotor systems in vocal emotional processing in children. *NeuroImage, 201,* 116052. http://dx.doi.org/10.1016/j.neuroimage.2019.116052

Cuddy, L. L., & Lyons, H. I. (1981). Musical pattern recognition: A comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology: Music, Mind, and Brain, 1,* 15–33. http://dx.doi.org/10.1037/h0094283

Dawson, C., Aalto, D., Šimko, J., Vainio, M., & Tervaniemi, M. (2017). Musical sophistication and the effect of complexity on auditory discrimination in Finnish speakers. *Frontiers in Neuroscience, 11,* 213. http://dx.doi.org/10.3389/fnins.2017.00213

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology, 2,* 124. http://dx.doi.org/10.3389/fpsyg.2011.00124

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion, 11,* 860–865. http://dx.doi.org/10.1037/a0022758

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex, 23,* 2038–2043. http://dx.doi.org/10.1093/cercor/bhs180

Frey, A., François, C., Chobert, J., Velay, J. L., Habib, M., & Besson, M. (2019). Music training positively influences the preattentive perception of voice onset time in children with dyslexia: A Longitudinal study. *Brain Sciences, 9,* 91. http://dx.doi.org/10.3390/brainsci9040091

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition and Emotion, 22,* 1094–1118. http://dx.doi.org/10.1080/02699930701626582

Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing, 38,* 455–464. http://dx.doi.org/10.1097/AUD.0000000000000402

Habibi, A., Cahn, B. R., Damasio, A., & Damasio, H. (2016). Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience, 21,* 1–14. http://dx.doi.org/10.1016/j.dcn.2016.04.003

Harrison, P., Musil, J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: Descriptive and explanatory approaches. *Journal of New Music Research, 45,* 265–280. http://dx.doi.org/10.1080/09298215.2016.1197953

Hayes, A. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron, 76,* 486–502. http://dx.doi.org/10.1016/j.neuron.2012.10.011

Honing, H. (2019). *The evolving animal orchestra: In search of what makes us musical*. Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/11839.001.0001

Isaacowitz, D. M., Löckenhoff, C. E., Lane, R. D., Wright, R., Sechrest, L., Riedel, R., & Costa, P. T. (2007). Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychology and Aging, 22,* 147–159. http://dx.doi.org/10.1037/0882-7974.22.1.147

Iversen, J. R., & Patel, A. D. (2008). The beat alignment test (BAT): surveying beat processing abilities in the general population. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th International Conference on Music Perception and Cognition* (pp. 465–468), Sapporo, Japan.

Jarosz, A., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving, 7,* 2–9. http://dx.doi.org/10.7771/1932-6246.1167

JASP Team. (2018). *JASP* (Version 0.9.2) [Computer software]. Retrieved from https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York, NY: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. http://dx.doi.org/10.1037/0033-2909.129.5.770

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience, 15,* 170–180. http://dx.doi.org/10.1038/nrn3666

Koelsch, S. (2015). Music-evoked emotions: Principles, brain correlates, and implications for therapy. *Annals of the New York Academy of Sciences, 1337,* 193–201. http://dx.doi.org/10.1111/nyas.12684

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11,* 599–605. http://dx.doi.org/10.1038/nrn2882

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., . . . Scott, S. K. (2018). Beatboxers and guitarists engage sensorimotor fields when listening to the instruments they can play. *Cerebral Cortex, 28,* 4063–4079. http://dx.doi.org/10.1093/cercor/bhy208

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports, 6,* 34911. http://dx.doi.org/10.1038/srep34911

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion, 11,* 1021–1031. http://dx.doi.org/10.1037/a0024521

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45,* 1234–1245. http://dx.doi.org/10.3758/s13428-013-0324-3

Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2018). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*. Advance online publication. http://dx.doi.org/10.1177/0305735618801997

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences, 39,* 527–542. http://dx.doi.org/10.1016/j.tins.2016.06.003

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., . . . Scott, S. K. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex, 25,* 4638–4650. http://dx.doi.org/10.1093/cercor/bhv134

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*. Stockholm, Sweden: Department of Neurosciences, Karolinska Hospital.

MacGregor, C., & Müllensiefen, D. (2019). The Musical Emotion Discrimination Task: A new measure for assessing the ability to discriminate emotions in music. *Frontiers in Psychology, 10,* 1955. http://dx.doi.org/10.3389/fpsyg.2019.01955

Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports, 9,* 10404. http://dx.doi.org/10.1038/s41598-019-46728-1

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences of the United States of America, 115,* 13129–13134. http://dx.doi.org/10.1073/pnas.1811793115

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19,* 1453–1463. http://dx.doi.org/10.1162/jocn.2007.19.9.1453

McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2015). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex, 25,* 246–257. http://dx.doi.org/10.1093/cercor/bht227

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22,* 1425–1433. http://dx.doi.org/10.1177/0956797611416999

Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research, 308,* 84–97. http://dx.doi.org/10.1016/j.heares.2013.09.012

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19,* 712–723. http://dx.doi.org/10.1093/cercor/bhn120

Moritz, C. E., Yampolsky, S., Papadelis, G., Thomson, J. M., & Wolf, M. (2012). Links between early rhythm skills, musical training, and phonological awareness. *Reading and Writing, 26,* 739–769. http://dx.doi.org/10.1007/s11145-012-9389-0

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science, 25,* 1795–1803. http://dx.doi.org/10.1177/0956797614541990

Mualem, O., & Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education, 33,* 413–425. http://dx.doi.org/10.1177/0255761415584292

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE, 9,* e89642. http://dx.doi.org/10.1371/journal.pone.0089642

Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., . . . Bao, Y. (2015). Sadness is unique: Neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience, 8,* 1049. http://dx.doi.org/10.3389/fnhum.2014.01049

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2,* 142. http://dx.doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98–108. http://dx.doi.org/10.1016/j.heares.2013.08.011

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111,* 14–25. http://dx.doi.org/10.1016/j.biopsycho.2015.08.008

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688–691. http://dx.doi.org/10.1038/nn1083

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review, 20,* 747–753. http://dx.doi.org/10.3758/s13423-013-0401-8

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, Ó. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language, 140,* 24–34. http://dx.doi.org/10.1016/j.bandl.2014.10.009

Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music, 42,* 284–298. http://dx.doi.org/10.1177/0305735612471239

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001

Sala, G., & Gobet, F. (2017a). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science, 26,* 515–520. http://dx.doi.org/10.1177/0963721417712760

Sala, G., & Gobet, F. (2017b). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review, 20,* 55–67. http://dx.doi.org/10.1016/j.edurev.2016.11.005

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15,* 511–514. http://dx.doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology, 102,* 283–302. http://dx.doi.org/10.1111/j.2044-8295.2010.02000.x

Schellenberg, E. G. (2019). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. http://dx.doi.org/10.1037/aca0000263

Schellenberg, E. G., Corrigall, K. A., Dys, S. P., & Malti, T. (2015). Group music training and children's prosocial skills. *PLoS ONE, 10,* e0141449. http://dx.doi.org/10.1371/journal.pone.0141449

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24–30. http://dx.doi.org/10.1016/j.tics.2005.11.009

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41,* 341–349. http://dx.doi.org/10.1111/1469-8986.00172.x

Scott, S. K., Sauter, D., & McGettigan, C. (2010). Brain mechanisms for processing perceived emotional vocalizations in humans. In S. M. Brudzynski (Ed.), *Handbook of behavioral neuroscience* (pp. 187–197). London, UK: Academic Press.

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition, 152,* 199–211. http://dx.doi.org/10.1016/j.cognition.2016.03.017

Soranzo, A., & Grassi, M. (2014). PSYCHOACOUSTICS: A comprehensive MATLAB toolbox for auditory testing. *Frontiers in Psychology, 5,* 712. http://dx.doi.org/10.3389/fpsyg.2014.00712

Strachan, J. W. A., Sebanz, N., & Knoblich, G. (2019). The role of emotion in the dyad inversion effect. *PLoS ONE, 14,* e0219185. http://dx.doi.org/10.1371/journal.pone.0219185

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Annals of the New York Academy of Sciences, 1169,* 209–213. http://dx.doi.org/10.1111/j.1749-6632.2009.04864.x

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review, 24,* 1929–1934. http://dx.doi.org/10.3758/s13423-017-1244-5

Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports, 8,* 9223. http://dx.doi.org/10.1038/s41598-018-27571-2

Swaminathan, S., & Schellenberg, E. G. (2019). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000798

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the association between music training and reading in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 992–999. http://dx.doi.org/10.1037/xlm0000493

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4,* 46–64. http://dx.doi.org/10.1037/1528-3542.4.1.46

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion, 8,* 838–849. http://dx.doi.org/10.1037/a0014080

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25,* 58–76. http://dx.doi.org/10.3758/s13423-017-1323-7

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25,* 35–57. http://dx.doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E. J., Verhagen, J., & Ly, A. (2016). How to quantify the evidence for the absence of a correlation. *Behavior Research Methods, 48,* 413–426. http://dx.doi.org/10.3758/s13428-015-0593-0

Wagner, H. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28. http://dx.doi.org/10.1007/BF00987006

Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J., . . . Scott, S. K. (2006). Positive emotions preferentially engage an auditory-motor "mirror" system. *The Journal of Neuroscience, 26,* 13067–13075. http://dx.doi.org/10.1523/JNEUROSCI.3907-06.2006

Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças-Third Edition (WISC III)*. Lisbon, Portugal: CEGOC-TEA.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10,* 420–422. http://dx.doi.org/10.1038/nn1872

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (2018). The definition of a musician in music psychology: A literature review and the six-year rule. *Psychology of Music*. Advance online publication. http://dx.doi.org/10.1177/0305735618804038

Received November 3, 2019
Revision received February 26, 2020
Accepted March 9, 2020 ■